



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년03월02일
(11) 등록번호 10-1711677
(24) 등록일자 2017년02월23일

(51) 국제특허분류(Int. Cl.)
H04N 5/92 (2006.01) G06K 9/00 (2006.01)
G06K 9/62 (2006.01)
(21) 출원번호 10-2011-7022868
(22) 출원일자(국제) 2010년04월01일
심사청구일자 2015년03월06일
(85) 번역문제출일자 2011년09월29일
(65) 공개번호 10-2012-0003447
(43) 공개일자 2012년01월10일
(86) 국제출원번호 PCT/US2010/029709
(87) 국제공개번호 WO 2010/115056
국제공개일자 2010년10월07일
(30) 우선권주장
12/416,152 2009년04월01일 미국(US)
(56) 선행기술조사문헌
Eric Galmar et al., "Representation and Analysis of Video Content for Automatic Object Extraction", EURECOM Research Report RR-10-231, 23 June 2008.*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
베이커 시몬 제이
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 인터내셔널 페이턴츠 마이크로소프트
코포레이션
지트닉 3세 찰스 로렌스
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 인터내셔널 페이턴츠 마이크로소프트
코포레이션
슈로프 게하드 플로리안
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 인터내셔널 페이턴츠 마이크로소프트
코포레이션
(74) 대리인
제일특허법인

전체 청구항 수 : 총 15 항

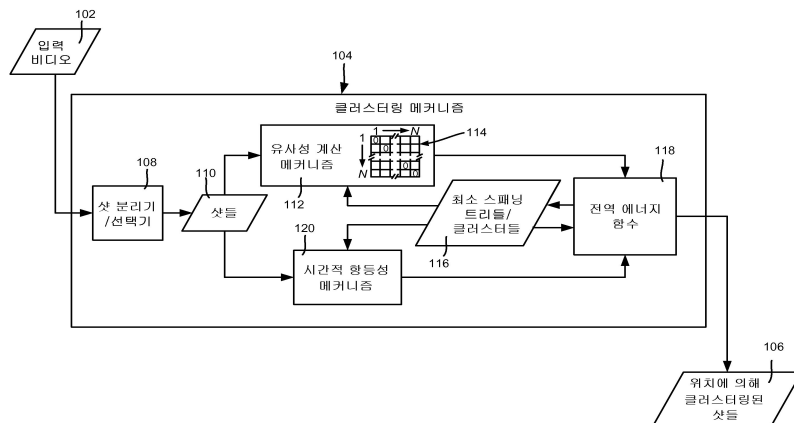
심사관 : 이성현

(54) 발명의 명칭 위치에 의한 비디오들의 클러스터링

(57) 요약

비디오 샷들을 캡처한 위치에 기초하여 샷들을 클러스터링하는 기술이 개시된다. 상당히 밀도가 높고 양호하게 연결되도록 클러스터들을 계산하는 제1 항을 포함하는 전역 에너지 함수를 최적화하여, 예컨대 샷들의 쌍들 사이의 유사성 스코어들에 기초하여, 한 위치에서 캡처 됐을 수 있는 샷들을 매칭시킨다. 제2 항은 동일 클러스터 내에 배치되도록 후속 샷들을 유도하는 시간적 사전 확률이다. 샷들은 샷들 각각의 노드들에 의해 표현되는 샷들 사이의 유사성 스코어에 기초하는 가중치들을 갖는 에지들을 구비하는 최소 스패닝 트리의 노드들로서 표현될 수 있다. 이용 가능한 클러스터들의 쌍들을 선택하고, 쌍들을 병합하고, 최저 비용을 갖는 쌍을 유지함으로써 응집 클러스터링이 수행된다. 중지 기준 또는 기준들이 충족될 때까지(예컨대, 단일 클러스터만이 남을 때까지) 클러스터들이 반복적으로 병합된다.

대표도



명세서

청구범위

청구항 1

컴퓨팅 환경에서, 복수의 샷을 포함하는 입력 비디오를 처리하는 방법으로서,
 샷들이 동일 위치에서 캡처되었는지를 나타내는 상기 샷들 간의 유사성을 결정하는 단계(306)와,
 상기 유사성을 전역 에너지 함수(global energy function)의 일부로서 이용하여 위치에 의해 샷들을 함께 클러스터링하는 단계(314)를 포함하는
 방법.

청구항 2

제1항에 있어서,
 상기 유사성을 전역 에너지 함수의 일부로서 이용하여 위치에 의해 샷들을 함께 클러스터링하는 단계는 샷들을 함께 클러스터링하는 비용(cost)을 나타내는 최소 스패닝 트리(minimum spanning trees)를 처리하는 단계를 포함하는
 방법.

청구항 3

제1항에 있어서,
 상기 전역 에너지 함수는 시간적 사전 확률 항(temporal prior term)을 포함하며, 상기 방법은 시간적 시퀀스 내에서 상이한 클러스터 내에 있는 이웃하는 샷들에 페널티를 부과하기 위해 상기 시간적 사전 확률 항을 적용하는 단계를 더 포함하는
 방법.

청구항 4

제1항에 있어서,
 상기 입력 비디오를 복수의 프레임 세트로 분리하는 단계와,
 각각의 프레임 세트로부터 적어도 하나의 키 프레임을 상기 세트를 나타내는 샷 또는 샷들로서 선택하는 단계를 더 포함하는
 방법.

청구항 5

제4항에 있어서,
 상기 세트의 상기 키 프레임은 상기 프레임 세트 내에서 시간적으로 중심에 있는 프레임을 포함하는
 방법.

청구항 6

제4항에 있어서,
 적어도 하나의 키 프레임을 선택하는 단계는 상기 프레임 세트로부터 복수의 키 프레임을 샘플링하는 단계를 포함하며,
 상기 방법은 상기 세트로부터 샘플링된 상기 복수의 키 프레임을 초기에 함께 클러스터링하는 단계를 더 포함하

는

방법.

청구항 7

제1항에 있어서,

상기 샷들 간의 상기 유사성을 결정하는 단계는 상기 샷들 각각에 대한 텍스톤 히스토그램(texton histogram)을 결정하는 단계를 포함하는

방법.

청구항 8

제1항에 있어서,

상기 샷들 간의 상기 유사성을 결정하는 단계는 상기 샷들 각각을 나타내는 벡터를 계산하는 단계를 포함하되, 상기 벡터는 상기 샷 내의 전경 정보보다 상기 샷 내의 배경 정보를 강조하는

방법.

청구항 9

제1항에 있어서,

상기 유사성을 전역 에너지 함수의 일부로서 이용하여 위치에 의해 샷들을 함께 클러스터링하는 단계는 클러스터들의 쌍들을 선택하는 단계, 각각의 쌍을 병합된 후보 클러스터로 병합하는 단계, 최저 비용을 갖는 상기 병합된 후보 클러스터를 유지하는 단계, 및 중지 기준 또는 기준들이 충족될 때까지 클러스터들을 더 병합하는 것을 반복하는 단계를 포함하는

방법.

청구항 10

컴퓨팅 환경에서, 샷들의 쌍들 간의 유사성 스코어에 기초하여 응집 클러스터링(agglomerative clustering)을 이용하여 전역 에너지 함수를 최적화함으로써, 비디오 프레임들을 나타내는 샷들을 유사한 위치를 갖는 샷들의 클러스터로 클러스터링하는 클러스터링 메커니즘부를 포함하는

장치.

청구항 11

제10항에 있어서,

상기 클러스터링 메커니즘부는 샷들 간의 시간적 일치성(temporal consistency)에 기초하여 상기 전역 에너지 함수를 더 최적화하는

장치.

청구항 12

제11항에 있어서,

상기 전역 에너지 함수는 유사성 스코어 데이터와 시간적 일치성 데이터의 합에 기초하며, 상기 유사성 스코어 데이터 및 상기 시간적 일치성 데이터가 서로 상대적으로 상기 합에 얼마나 많이 기여하는지를 제어하기 위해 가중 팩터가 사용되는

장치.

청구항 13

제10항에 있어서,

상기 클러스터링 메커니즘부는 제각기의 노드에 의해 표현되는 상기 샷들 간의 상기 유사성 스코어에 적어도 부분적으로 기초하는 가중치를 갖는 에지를 구비하는 최소 스패닝 트리의 노드로서 상기 샷들을 배열하는 장치.

청구항 14

실행될 때,

샷 경계 검출에 기초하여 비디오를 프레임 세트로 분리하는 단계(302)와,

각각의 프레임 세트로부터 적어도 하나의 키 프레임을 선택하는 단계(304)와,

각각의 세트의 키 프레임 또는 키 프레임들 간의 유사성에 기초하여 유사성 스코어를 계산하는 단계(306)와,

키 프레임이 다른 키 프레임과 시간적으로 일치하는지에 기초하여 시간 데이터를 계산하는 단계(314)와,

상기 유사성 스코어 및 상기 시간 데이터를 이용하여 자신들의 키 프레임에 의해 표현되는 샷들을 함께 클러스터링하는 단계(314)

를 실행하는 컴퓨터 실행가능 명령어를 구비하는 하나 이상의 하드웨어 컴퓨터 저장 장치.

청구항 15

제14항에 있어서,

키 프레임들의 쌍에 대한 상기 유사성 스코어 및 상기 시간 데이터는 비용에 대응하고, 상기 유사성 스코어 및 상기 시간 데이터를 이용하여 샷들을 클러스터링하는 단계는 클러스터들의 쌍들을 선택하는 단계, 각각의 쌍을 병합된 후보 클러스터로 병합하는 단계, 최저 비용을 갖는 상기 병합된 후보 클러스터를 유지하는 단계, 및 중지 기준 또는 기준들이 충족될 때까지 클러스터를 더 병합하는 것을 반복하는 단계를 포함하는

하나 이상의 하드웨어 컴퓨터 저장 장치.

발명의 설명

배경 기술

[0001] 이를테면 소정의 원하는 세그먼트들을 선택하기 위해 비디오를 볼 때, 위치는 다양한 작업들에 유용한 정보 소스일 수 있다. 예컨대, 사용자는 아이의 놀이방에서의 홈 비디오 샷이 사용자가 친지에게 전송하기를 원하는 특정한 장면을 포함한다는 것을 생각할 수 있으며, 이에 따라 그 위치에서 촬영된 비디오들의 비디오 세그먼트들(또는 대표적인 이미지들)을 빨리 찾는 것이 유용할 것이다. 일반적으로, 사용자들은 위치에 의해 비디오들을 브라우징 또는 검색하고, 위치들에 주석을 달며/달거나 위치에 특정한 편집물들을 생성하기를 원할 수 있다.

[0002] 위치 기반 클러스터링 알고리즘들은 그러한 작업에서 사용자들을 지원하려고 시도한다. 그러나, 위치 기반 클러스터링 알고리즘들에 대한 하나의 중요한 도전은 단일 위치 내에 나타나는 광범위한 모습들(appearances)이다. 예를 들어, 집의 동일한 방 안에서 촬영된 비디오를 고려해 보라. 각각의 샷이 캡처된 위치에 대한 관점에 따라, 매우 다른 모습들이 가능하다.

발명의 내용

과제의 해결 수단

[0003] 본 요약은 아래의 상세한 설명에서 더 설명되는 대표적인 개념들의 발체를 간단하게 소개하기 위해 제공된다. 본 요약은 청구 발명의 주요 특징들 또는 본질적인 특징들을 식별하는 것을 의도하지 않으며, 어떠한 식으로도 청구 발명의 범위를 한정하는 데 사용하고자 하는 의도도 없다.

[0004] 간단히 말하면, 본 명세서에서 설명되는 본 발명의 다양한 양태들은, 응집 클러스터링을 이용하여 전역 에너지 함수를 최적화함에 의한 것을 포함하여, 유사한 위치들을 갖는 것에 기초하여, 비디오 프레임들을 나타내는 샷들을 클러스터링하는 기술에 관한 것이다. 샷들의 쌍들 사이의 유사성 점수들은 물론, 샷들 사이의 시간적 일

관성을 나타내는 값이 계산된다. 전역 에너지 함수는 샷들에 대한 유사성 스코어 데이터와 (상대적인 기여에 대해 가중된) 시간적 일관성 데이터를 합산한다. 일 구현에서, 샷들은 샷들 각각의 노드들에 의해 표현되는 샷들 사이의 유사성 스코어에 적어도 부분적으로 기초하는 가중치들을 갖는 에지들을 구비한 최소 스패닝 트리(minimum spanning tree)의 노드들에 의해 표현된다.

[0005] 샷을 나타내는 하나의 키 프레임(또는 동일 샷으로부터 취해진 다수의 키 프레임)을 갖도록 클러스터들을 초기화하고 클러스터들의 쌍들을 선택함으로써 응집 클러스터링이 수행된다. 각각의 쌍이 후보 클러스터 내에 병합되어, 병합된 후보 클러스터 중 최저의 비용을 가진 클러스터가 새로운 클러스터로 유지된다. 중지 기준 또는 기준들이 충족될 때까지(예컨대, 소정 수의 클러스터들만이 남을 때까지) 클러스터들이 반복적으로 병합된다.

[0006] 도면들과 관련하여 고려될 때 아래의 상세한 설명으로부터 다른 이점들이 명확해질 수 있다.

도면의 간단한 설명

[0007] 본 발명은 동일한 참조 번호들이 유사한 요소들을 지시하는 첨부 도면들에 한정이 아니라 예시적으로 도시된다.

도 1은 위치에 기초하여 비디오들을 클러스터링하기 위한 예시적인 컴포넌트들을 나타내는 블록도이다.

도 2는 에너지 함수를 통해 모델링된 중간 클러스터들을 나타낸다.

도 3은 위치에 기초하여 비디오들을 클러스터링하기 위한 예시적인 단계들을 나타내는 흐름도이다.

도 4는 본 발명의 다양한 양태들이 통합될 수 있는 컴퓨팅 환경의 일례를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0008] 본 명세서에서 설명되는 기술의 다양한 양태들은 일반적으로 클러스터 비용(데이터 항) 및 시간적 사전 확률(temporal prior)을 포함하는 전역 에너지 함수를 최적화함에 의한 것을 포함하는 위치에 의한 비디오들의 클러스터링에 관한 것이다. 일 양태에서, 샷들의 표현 및 샷들 사이의 거리 척도(distance measure)가 결정된 후에 발생하는 클러스터링이 본 명세서에서 설명되는 바와 같이 최적화된다.

[0009] 본 명세서에서 설명되는 바와 같은 위치에 의한 클러스터링은 "홈" 비디오에 대해 양호하게 이루어지지만, 이것은 비디오의 일 타입일 뿐이라는 것을 이해한다. 텔레비전 또는 영화들에 나오는 것과 같은 전문적으로 캡처된 비디오도 위치에 의한 클러스터링이 언제 그리고 어디서(예컨대, 스튜디오에서 또는 시청자의 개인 소장품에서) 이루어지는지에 관계없이 본 명세서에서 설명되는 기술로부터 이익을 얻을 수 있다. 따라서, 본 발명은 본 명세서에서 설명되는 임의의 특정 실시예들, 양태들, 개념들, 구조들, 기능들 또는 예들로 한정되지 않는다. 오히려, 본 명세서에서 설명되는 실시예들, 양태들, 개념들, 구조들, 기능들 또는 예들 중 어느 것도 한정적이지 아니며, 본 발명은 일반적으로 컴퓨팅 및 비디오 처리에서 이익들 및 이점들을 제공하는 다양한 방식으로 이용될 수 있다.

[0010] 도 1은 위치에 의한 비디오들의 클러스터링과 관련된 다양한 양태들을 나타낸다. 일반적으로, 임의의 적절한 소스로부터의 입력 비디오(102)는 클러스터링 메커니즘(104)에 의해 처리되어, 샷들의 클러스터들(106)이 된다. 이를 위해, 샷 분리기(108)가 비디오(102)를 샷들(110)로 분리하며, 각각의 샷은 하나 이상의 프레임들을 포함한다. 일반적으로, 샷 경계들은 카메라가 오프 상태에서 온 상태로 바뀌거나 캡처하고 있는 것을 빠르게 바꿀 때에 결정되며, 이는 공지된 기술들을 포함하는 임의의 적절한 검출 수단에 의해 이루어질 수 있는데, 예컨대 마이크로소프트 무비 메이커(Microsoft[®] Movie Maker)가 그러한 기능을 제공한다.

[0011] 샷들(110)로 분리된 후, 도 1에 도시된 바와 같이, 유사성 계산 메커니즘(112)에 의해 샷들 각각 사이의 유사성 스코어가 계산된다. 일 구현에서, 샷들 사이의 유사성은 후술하는 페어와이즈 거리 함수(pair-wise distance function)에 의해 결정된다.

[0012] 하나의 대안에서는, 샷을 구성하는 프레임들의 시리즈 내의 각각의 프레임에 대한 스코어를 계산하는 것이 아니라, 샷의 중심 프레임이 유사성 비교를 위한 키 프레임으로서 선택될 수 있다. 다른 대안으로서, 일정한 샘플링 레이트로, 예컨대 10 프레임마다 샘플링함으로써 샷의 소정의 감소된 수의 프레임들이 다수의 키 프레임으로서 선택될 수 있다(혹은, 소정의 최소 수의 프레임들보다 적은 경우에는 중심 프레임이 사용될 수 있다). 후술하는 바와 같이, 단일 샷의 다수의 키 프레임은 자동으로 함께 클러스터링될 수 있다. 다양한 프레임들에 기초하여 모자이크를 계산하는 것이 가능하지만, 줌들 및 (예컨대, 사람의) 움직임은 이러한 접근법에서의 어려움을 제공한다.

- [0013] 유사성에 대하여 비교될 수도 있는 키 프레임들에 대한 수치적 표현들을 설정하기 위하여, 일 구현은 무작위로 샘플링된 5x5 패치들(patches) 및 k-평균(k-means) 클러스터링을 이용하여 오프라인으로 생성되는 128개의 텍스톤(texton)을 갖는 것과 같은 텍스톤 어휘에 대한 평가를 위해 텍스톤들의 히스토그램을 계산한다. 구체적으로, 샷 내의 각각의 키 프레임에 대해, 고밀도 그리드 내에서 5x5 패치들이 추출된다. 이어서, 각각의 패치는 가장 가까운 텍스톤에 할당되며, 전체 키 프레임에 걸쳐 집계함으로써 텍스톤들에 대한 히스토그램이 생성된다. 한 쌍의 텍스톤 히스토그램들 사이의 거리를 계산하고, 따라서 프레임들 사이의 유사성을 결정하기 위하여, 공지된 카이-제곱 거리 계산(Chi-Squared distance computation)이 이용될 수 있다. 잠재적 디리클레 할당(Latent Dirichlet Allocation)을 이용하는 것이 하나의 대안이라는 점에 유의한다.
- [0014] 유사성을 결정하기 위하여 대안적인 메커니즘들이 이용될 수 있다는 점에 유의한다. 예컨대, 하나의 대안적인 키프레임간(inter-keyframe) 거리 함수는 특징 기반이다. 유사한 불변 특징들(affine invariant features)의 세트가 공지 기술들을 이용하여 발견된다. (예컨대, 백만 개의 리프(leaf) 노드를 갖는) 어휘 트리를 이용하여 특징들에 의해 추출된 각각의 이미지 패치에 비주얼 워드(visual word)가 할당된다. 이미지들 사이의 유사성 스코어는 공지된 용어 빈도 - 도치 문서 빈도(TF-IDF) 점수 개념을 이용하여 발견된다.
- [0015] N개의 키 프레임에 대한 유사성 스코어들은 기본적으로 도 1에 도시된 바와 같은 스코어들의 그리드(114)를 형성한다. 후술하는 바와 같이, 이러한 스코어들로부터 클러스터들이 생성되어 최소 스패닝 트리들(116)의 세트로서 배열된다. 일반적으로, 전역 에너지 함수(118)가 이러한 최소 스패닝 트리들(116)을 처리하여 샷들의 클러스터들(106)을 얻는다.
- [0016] 일 구현에서, 전역 에너지 함수(118)는 클러스터 비용(또는 데이터) 항 및 시간적 사전 확률 항을 포함하는 2개의 항을 포함한다.

수학식 1

[0017]
$$E_{Global} = E_{Cluster} + \lambda E_{Temporal} \quad (1)$$

- [0018] 여기서, $E_{Cluster}$ 는 클러스터들의 모델(데이터 항)이고, $E_{Temporal}$ 은 시간적 사전 확률 항이며, λ 는 가중치이다.
- [0019] 클러스터 모델에서, 동일 위치에서 캡처된 샷들의 세트는 아마도 특유의 구조를 가질 것이다. 샷들의 일부 쌍들은 근본적으로 상이한 뷰포인트들을 가질 수 있는 반면, 다른 쌍들은 매우 유사할 수 있지만, 일반적으로 클러스터는 상당히 밀도가 높고, 양호하게 연결될 것으로 직관적으로 예상된다. 일반적으로, 이러한 개념들은 클러스터 비용($E_{Cluster}$) 내에 포함된다.
- [0020] 클러스터링 메커니즘은 공지된 응집 클러스터링 기술들을 이용하여 수학식 1 내의 전역 에너지(E_{Global})를 최적화한다. 일반적으로, 응집 클러스터링은 처음에 각각의 샷을 그 자신의 클러스터에 할당하지만, 샷당 둘 이상의 키 프레임을 허용하는 대안에서는 동일 샷의 프레임들이 처음에 동일 클러스터의 일부가 되도록 클러스터링 알고리즘이 초기화된다는 점에 유의한다.
- [0021] 초기화되면, 클러스터들의 쌍들이 반복적으로 병합된다. 각각의 반복에서, 2개의 클러스터(C_i, C_j)의 모든 가능한 병합의 전역 비용에 대한 변화가 고려되며, 최저 전역 비용(E_{Global})을 발생시키는 병합이 이용된다. 이것은 샷들을 포함하는 단일 클러스터만이 남을 때(또는 10개의 클러스터와 같은 소정의 다른 수의 클러스터가 남을 때) 완료된다. 즉, 비주얼 데이터의 구조를 모델링하는 클러스터 비용/데이터 항은 k-1개의 최소 스패닝 트리를 제거하여 k 번째 스패닝 트리를 계산한 후에 클러스터의 모델이 k 번째 최소 스패닝 트리의 총 길이를 갖는 다수의 최소 스패닝 트리를 계산함으로써 얻어질 수 있다(여기서, k는 샷들의 수에 비례한다).
- [0022] 결과적으로, 비주얼 데이터에 좀처럼 적용되지 않는 길고 얇은 클러스터 또는 소형 클러스터 대신에, (도 2의 중간 클러스터와 같은) 중간 클러스터들의 모델이 중간 클러스터들에 적합한 에너지 함수의 형태로 사용된다.

수학식 2

$$E_{\text{Cluster}} = \sum_i \text{MST}(C_i^k) \quad (2)$$

[0023]

[0024] 여기서, MST는 최소 스패닝 트리를 나타내며, $k = \alpha(|C_i|-1)$ 은 임의의 주어진 노드의 이웃 노드들의 수의 분수 ($\alpha \in [0,1]$)이다. 여기서:

수학식 3

$$C_i^k = C_i^{k-1} - \text{MST}(C_i^{k-1}), \quad C_i^1 = C_i \quad (3)$$

[0025]

[0026] 위의 식은 C_i^{k-1} 로부터 MST 내의 에지들을 제거함으로써 C_i^{k-1} 를 계산할 수 있는 순환적 정의인데, 즉 C_i^k 는 C_i 로부터 시퀀스 내의 k-1개의 MST를 제거한 후에 얻어진 그래프이다. 그래프가 끊어질 가능성을 피하기 위하여, 에지들을 제거하는 대신에, 에지들을 샷 매치 스코어의 최대값으로 대체할 수 있다는 점에 유의한다. 전술한 바와 같이, 2개의 샷 사이의 페어와이즈 거리 함수 $d(s_{t1}, s_{t2})$ 는 클러스터 C_i 내에서의 이들 사이의 매칭 비용, 즉 시간 $t1$ 과 $t2$ 에서의 2개의 샷 s_{t1} 과 s_{t2} 의 벡터 표현들 사이의 거리이다.

[0027]

파라미터 α 에 대한 값의 선택은 클러스터들이 얼마나 길고 얇을 수 있는지에 대한 제어를 제공하며, 이에 따라 α 의 적절한 값에 대해, 수학식 2 내의 클러스터 비용은 길고 얇은 클러스터들을 불허하지만 중간 클러스터들을 허용한다. 클러스터들은 비교적 길 것으로 예상되는데, 그 이유는 방의 상이한 부분들의 모습이 아주 다를 수 있기 때문이다. 이와 동시에, 가능한 카메라 뷰포인트들 및 유사한(호감이 가는) 뷰포인트들을 갖는 다수의 샷의 연속체가 예상되며, 이에 따라 클러스터는 또한 아주 밀도가 높을 것으로 예상된다. 일 구현에서는, $\alpha = 0.3$ 의 중간 값이 사용된다.

[0028]

시간적 사전 확률 항(가중 인자로서 λ 를 갖는 E_{Temporal})과 관련하여, 후속 샷들은 동일 위치를 표시할 가능성이 더 크다. 일반적으로, 시간적 사전 확률 항은 비디오 내의 2개의 후속 샷이 동일 위치가 아닌 것보다 동일 위치에서 캡처될 가능성이 더 클 가능성에 기초한다. 따라서, 비디오로부터의 시간 시퀀스 내의 이웃 샷들의 각각의 쌍이 고려되며, 이 경우에 상이한 클러스터들 내에 있는 각각의 쌍에 대해 (예컨대, 도 1의 시간적 일치성 메커니즘에 의해) 페널티가 부과되고, 그렇지 않은 경우에는 추가되지 않는다.

[0029]

이전에는 비디오를 동일 위치에서 캡처된 장면들로 분할하기 위해 하드 시간 세그먼트화(hard temporal segmentation)가 수행되었지만, 홈 비디오에서의 감소된 시간적 일치성은 그러한 하드 판정을 부적절하게 하며, 그 대신에 소프트 시간 사전 확률을 이용하여 결과들에서 더 높은 시간적 일치성을 제공한다.

수학식 4

$$E_{\text{Temporal}} = \sum_t \delta(s_t, s_{t+1}) \quad (4)$$

[0030]

[0031] 여기서, $\delta(s_t, s_{t+1})$ 은 샷들(s_t, s_{t+1})이 상이한 클러스터들 내에 있는지를 결정하는 지시자 함수이다.

수학식 5

$$\delta(s_t, s_{t+1}) = \begin{cases} 1 & s_t \in C_i, s_{t+1} \in C_j, i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

[0032]

[0033] 수학식 4 및 5는 그러한 시간적으로 이웃하는 샷들이 상이한 클러스터들에 속하는 횟수를 셈한다. 마르코프(Markov) 랜덤 필드 용어에서, 이러한 수학식들은 이웃하는 프레임들이 상이한 위치에서 발생하는 경우에 이들 사이에 페널티가 추가되는 포츠 모델(Potts model)을 기술한다.

[0034]

다른 양태를 참조하면, 위치에 의한 클러스터링의 또 하나의 어려움은 일시적인 전경 물체들, 주로 위치 내에 때때로 나타나고, 동일 샷 내의 위치 주위를 때때로 돌아다니는 사람들의 존재이다. 게다가, 동일한 의복을 입은 동일한 사람들이 상이한 위치들에 종종 나타나서, 텍스톤/(잠재적 디리클레 할당) 토폭 거리들 및 특징 거리들 모두에 방해물들(distractors)을 추가한다. 이러한 이벤트들은 홈 비디오들 및 전문적으로 편집된 콘텐츠 모두에서 발생한다.

[0035]

일 구현에서는, 가우스 사전 확률(가우스 검출기의 차이)을 이용하여, 전경보다 배경일 가능성이 더 큰 이미지의 부분들에 추가적 가중치를 제공할 수 있다. 이러한 방식으로, 예컨대, 히스토그램들이 가우스 공간 사전 확률로 가중될 수 있다. 일반적으로, 사람들은 주로 이미지의 중심에 위치하며, 따라서 중심은 다른 부분들보다 더 작은 가중치를 부여받을 수 있다.

[0036]

요약으로서, 도 3은 비디오를 처리하여 위치 기반 클러스터들을 생성하는 예시적인 단계들을 나타내는 흐름도이다. 단계 302는 비디오를 샷들로 분리하는 단계를 나타내며, 단계 304는 각각의 샷에 대한 하나 이상의 키 프레임들의 선택을 나타낸다.

[0037]

단계 306은 유사한 샷들을 함께 클러스터링하는 데 사용하기 위한 유사성 스코어들을 계산하는 단계를 나타낸다. (과선 박스로서 선택사항인 것으로서 도시된) 단계 308은 가중치들을 조정하여 배경 위치를 강조하고, 이에 따라 샷들 내의 사람들과 같은 전경 물체들을 감안하는 단계를 나타낸다.

[0038]

단계 310-312는 클러스터들의 초기화와 관련된다. 일반적으로, 프레임들의 세트 중 하나의 프레임이 사용되는 경우에 처음에 클러스터당 하나의 키 프레임이 존재하거나(단계 311), 클러스터당 하나 이상의 키 프레임들이 존재하며, 따라서 동일 샷으로부터의 프레임들이 함께 클러스터링된다(단계 312).

[0039]

단계 314는 전술한 바와 같은 클러스터들의 병합을 나타낸다. 일반적으로, 클러스터들의 세트가 주어질 때, 클러스터들의 각각의 가능한 쌍은 병합된 후보 클러스터로서 병합되고, 각각의 후보에 대해 (임의의 시간적 사전 확률 페널티를 포함하는) 병합 비용이 계산된다. 최저 비용을 갖는 병합된 후보 클러스터가 유지된다.

[0040]

이어서, 단계 316은 소정의 중지 기준 또는 기준들이 충족될 때까지 반복적으로 루프 백(loop back)하여 클러스터들의 수를 줄인다. 하나의 예시적인 중지 기준은, 소정 개수(한 개까지 적게)의 클러스터가 남을 때까지 병합하고, 따라서 이어서 사용자가 각각의 클러스터를 브라우징하여 원하는 샷들의 세트를 발견할 수 있는 것을 포함한다. 예컨대, 사용자는 특정 방 안에서 촬영된 비디오들을 빠르게 찾을 수 있다. 사용자는 이러한 중지 개수를 변경할 수 있는데, 예컨대 샷들이 원하는 대로 클러스터링되지 않은 경우에 되돌아 가서 클러스터들의 총 수를 증가 또는 감소시킬 수 있다. 또 하나의 예시적인 기준은 소정의 비용이 달성됨에 기초할 수 있다.

[0041]

단계 318은 클러스터링된 샷들을 출력한다. 이것은 각각의 클러스터에 대한 대표 이미지 또는 이미지들의 세트일 수 있거나, 비디오의 일부 또는 전부일 수 있다. 각각의 키 프레임은 애플리케이션으로 하여금 함께 클러스터링되는 키 프레임들에 대응하는 소스 비디오 내의 샷들을 빨리 찾을 수 있게 하는 적절한 식별자 등을 가질 수 있다.

[0042]

예시적인 운영 환경

[0043]

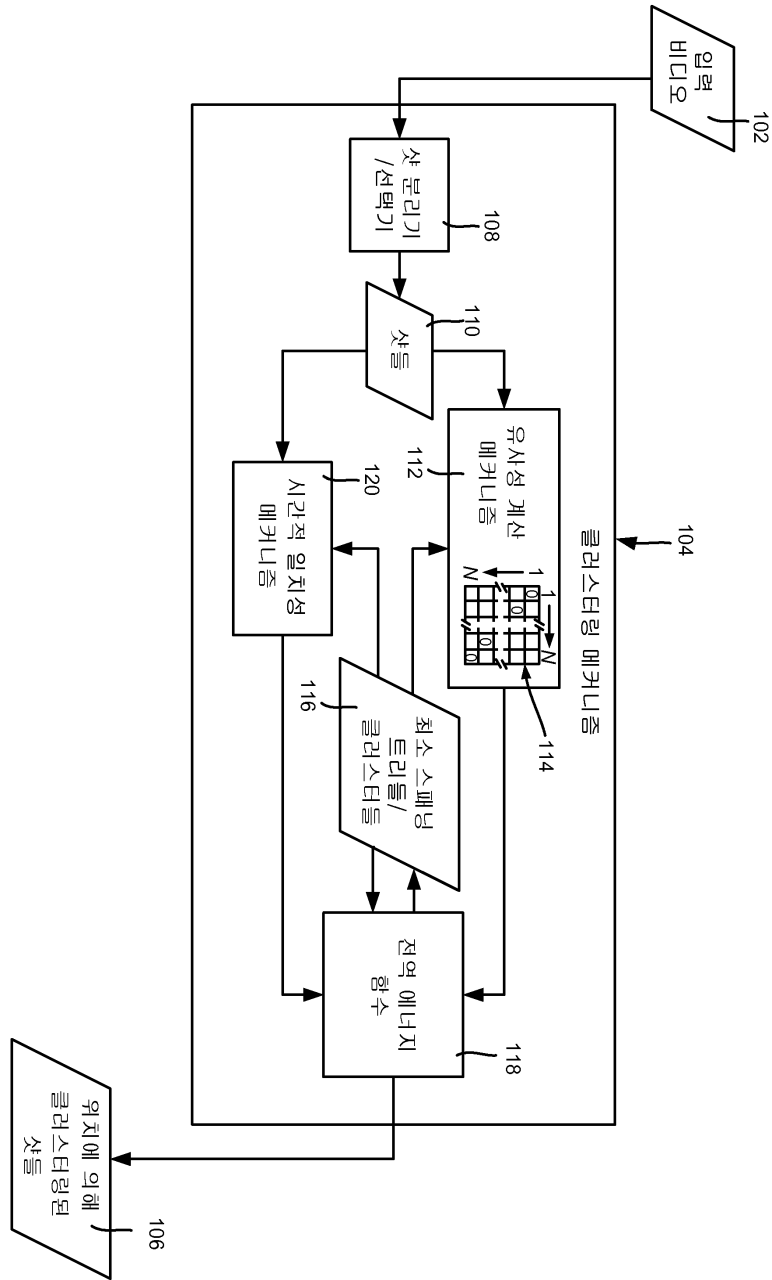
도 4는 도 1-3의 예들을 구현할 수 있는 적절한 컴퓨팅 및 네트워킹 환경(400)의 일례를 나타낸다. 컴퓨팅 시스템 환경(400)은 적절한 컴퓨팅 환경의 일례일 뿐이며, 본 발명의 이용 또는 기능의 범위에 대하여 어떠한 제한도 제시하는 것을 의도하지 않는다. 컴퓨팅 환경(400)은 예시적인 운영 환경(400) 내에 도시된 컴포넌트들 중 어느 하나 또는 조합에 관하여 어떠한 의존성 또는 요구도 갖지 않는 것으로 해석되어야 한다.

- [0044] 본 발명은 다양한 다른 범용 또는 특수 목적 컴퓨팅 환경 또는 구성에서 동작한다. 본 발명과 함께 사용하기에 적합할 수 있는 잘 알려진 컴퓨팅 시스템, 환경 및/또는 구성의 예는 개인용 컴퓨터, 서버 컴퓨터, 핸드헬드 또는 랩탑 장치, 태블릿 장치, 멀티프로세서 시스템, 마이크로프로세서-기반 시스템, 셋톱 박스, 프로그래밍 가능한 소비자 전자 장치, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 전술한 시스템들 또는 장치들 중의 임의의 것을 포함하는 분산 컴퓨팅 환경 등을 포함하지만 이에 한정되지 않는다.
- [0045] 본 발명은 컴퓨터에 의해 실행되는 프로그램 모듈 등의 컴퓨터 실행가능 명령어의 통상적 맥락에서 설명될 수 있다. 일반적으로, 프로그램 모듈은 특정 작업을 수행하거나 특정 추상 데이터 타입을 구현하는 루틴, 프로그램, 객체, 컴포넌트, 데이터 구조 등을 포함한다. 본 발명은 또한 통신 네트워크를 통해 링크된 원격 처리 장치에 의해 작업들이 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈들은 메모리 저장 장치를 포함하는 로컬 및/또는 원격 컴퓨터 저장 매체에 배치될 수 있다.
- [0046] 도 4를 참조하면, 본 발명의 다양한 양태들을 구현하기 위한 예시적인 시스템은 컴퓨터(410)의 형태의 범용 컴퓨팅 장치를 포함할 수 있다. 컴퓨터(410)의 컴포넌트들은 중앙 처리 유닛(420), 시스템 메모리(430), 및 시스템 메모리를 포함하는 다양한 시스템 컴포넌트를 처리 유닛(420)에 결합하는 시스템 버스(421)를 포함할 수 있지만, 이에 한정되는 것은 아니다. 시스템 버스(421)는 다양한 버스 아키텍처 중 임의의 것을 사용하는 메모리 버스 또는 메모리 컨트롤러, 주변 장치 버스 및 로컬 버스를 포함하는 여러 유형의 버스 구조들 중 임의의 것일 수 있다. 제한이 아니라 예로서, 이러한 아키텍처들은 산업 표준 아키텍처(ISA) 버스, 마이크로 채널 아키텍처(MCA) 버스, 향상된 ISA(EISA) 버스, 비디오 전자공학 표준 협회(VESA) 로컬 버스, 및 메자닌 버스(Mezzanine bus)로도 알려진 주변 컴포넌트 상호접속(PCI) 버스를 포함한다.
- [0047] 컴퓨터(410)는 일반적으로 다양한 컴퓨터 관독가능 매체를 포함한다. 컴퓨터 관독가능 매체는 컴퓨터(410)에 의해 액세스될 수 있고 휘발성 및 비휘발성, 이동식 및 비이동식 매체 모두를 포함하는 임의의 이용 가능한 매체일 수 있다. 제한이 아니라 예로서, 컴퓨터 관독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 관독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터 등의 정보를 저장하기 위한 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, 디지털 다기능 디스크(DVD) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 원하는 정보를 저장하는 데 사용될 수 있고 컴퓨터(410)에 의해 액세스될 수 있는 임의의 다른 매체를 포함하지만, 이에 한정되지 않는다. 통신 매체는 통상적으로 반송파 또는 기타 전송 매커니즘 등의 변조된 데이터 신호 내에 컴퓨터 관독 가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터를 구현하며 임의의 정보 전달 매체를 포함한다. "변조된 데이터 신호"라는 용어는 신호 내에 정보를 인코딩하는 방식으로 신호의 특성들 중 하나 이상이 설정 또는 변경된 신호를 의미한다. 제한이 아니라 예로서, 통신 매체는 유선 네트워크 또는 직접 유선 접속 등의 유선 매체와, 음향, RF, 적외선 및 기타 무선 매체 등의 무선 매체를 포함한다. 상술한 것들 중 임의의 것들의 조합들도 컴퓨터 관독 가능 매체의 범위 내에 포함될 수 있다.
- [0048] 시스템 메모리(430)는 관독 전용 메모리(ROM)(431) 및 랜덤 액세스 메모리(RAM)(432) 등의 휘발성 및/또는 비휘발성 메모리의 형태의 컴퓨터 저장 매체를 포함한다. 이를테면 시동 중에 컴퓨터(410) 내의 구성 요소들 간에 정보를 전송하는 것을 돕는 기본 루틴을 포함하는 기본 입출력 시스템(433)(BIOS)은 일반적으로 ROM(431)에 저장된다. RAM(432)은 일반적으로 처리 유닛(420)에 의해 즉시 액세스될 수 있고 그리고/또는 현재 조작되고 있는 데이터 및/또는 프로그램 모듈을 포함한다. 제한이 아니라 예로서, 도 4는 운영 체제(434), 애플리케이션 프로그램(435), 기타 프로그램 모듈(436) 및 프로그램 데이터(437)를 도시한다.
- [0049] 컴퓨터(410)는 또한 다른 이동식/비이동식, 휘발성/비휘발성 컴퓨터 저장 매체를 포함할 수 있다. 단지 예로서, 도 4는 비이동식 비휘발성 자기 매체를 관독하거나 그 자기 매체에 기록하는 하드 디스크 드라이브(441), 이동식 비휘발성 자기 디스크(452)를 관독하거나 그 자기 디스크에 기록하는 자기 디스크 드라이브(451), 및 CD-ROM 또는 기타 광학 매체 등의 이동식 비휘발성 광 디스크(456)를 관독하거나 그 광 디스크에 기록하는 광 디스크 드라이브(455)를 도시하고 있다. 예시적인 운영 환경에서 사용될 수 있는 다른 이동식/비이동식, 휘발성/비휘발성 컴퓨터 저장 매체는 자기 테이프 카세트, 플래시 메모리 카드, 디지털 다기능 디스크, 디지털 비디오 테이프, 반도체 RAM, 반도체 ROM 등을 포함하지만, 이에 한정되지 않는다. 하드 디스크 드라이브(441)는 일반적으로 인터페이스(440)와 같은 비이동식 메모리 인터페이스를 통해 시스템 버스(421)에 연결되고, 자기 디스크 드라이브(451) 및 광 디스크 드라이브(455)는 일반적으로 인터페이스(450)와 같은 이동식 메모리 인터페이스에 의해 시스템 버스(421)에 연결된다.

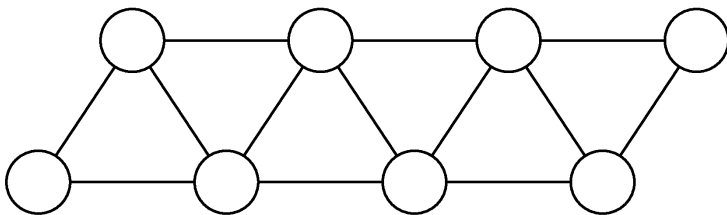
- [0050] 위에 설명되고, 도 4에 도시된 드라이브들 및 그들과 관련된 컴퓨터 저장 매체들은 컴퓨터(410)를 위한 컴퓨터 관독가능 명령어, 데이터 구조, 프로그램 모듈 및 기타 데이터의 저장을 제공한다. 도 4에서, 예를 들어, 하드 디스크 드라이브(441)는 운영 체제(444), 애플리케이션 프로그램(445), 기타 프로그램 모듈(446) 및 프로그램 데이터(447)를 저장하는 것으로 도시된다. 이들 컴포넌트는 운영 체제(434), 애플리케이션 프로그램(435), 기타 프로그램 모듈(436) 및 프로그램 데이터(437)와 동일할 수도 있고 다를 수도 있다. 운영 체제(444), 애플리케이션 프로그램(445), 기타 프로그램 모듈(446) 및 프로그램 데이터(447)는 최소한 이들이 상이한 사본들임을 나타내기 위하여 다른 번호들을 부여받고 있다. 사용자는 태블릿 또는 전자 디지털izer(464), 마이크로폰(463), 키보드(462), 및 일반적으로 마우스, 트랙볼 또는 터치 패드로서 지칭되는 포인팅 장치(461)와 같은 입력 장치들을 통해 컴퓨터(410)에 명령 및 정보를 입력할 수 있다. 도 4에 도시되지 않은 기타 입력 장치는 조이스틱, 게임 패드, 위성 안테나, 스캐너 등을 포함할 수 있다. 이들 및 다른 입력 장치들은 흔히 시스템 버스에 연결된 사용자 입력 인터페이스(460)를 통해 처리 유닛(420)에 연결되지만, 병렬 포트, 게임 포트 또는 유니버설 직렬 포트(USB)와 같은 기타 인터페이스 및 버스 구조에 의해 연결될 수 있다. 모니터(491) 또는 다른 타입의 디스플레이 장치도 비디오 인터페이스(490) 등의 인터페이스를 통해 시스템 버스(421)에 연결된다. 모니터(491)는 터치 스크린 패널 등과 통합될 수도 있다. 모니터 및/또는 터치 스크린 패널은 태블릿 타입의 개인용 컴퓨터에서와 같이 컴퓨팅 장치(410)를 넣는 하우징에 물리적으로 결합될 수 있다는 점에 유의한다. 게다가, 컴퓨팅 장치(410)와 같은 컴퓨터는 출력 주변 장치 인터페이스(494) 등을 통해 연결될 수 있는 스피커(495) 및 프린터(496) 등의 기타 주변 출력 장치도 포함할 수 있다.
- [0051] 컴퓨터(410)는 원격 컴퓨터(480)와 같은 하나 이상의 원격 컴퓨터에 대한 논리적 접속을 이용하여 네트워킹된 환경에서 동작할 수 있다. 원격 컴퓨터(480)는 개인용 컴퓨터, 서버, 라우터, 네트워크 PC, 피어(peer) 장치 또는 기타 공통 네트워크 노드일 수 있으며, 도 4에는 메모리 저장 장치(481)만이 도시되었지만, 통상적으로는 컴퓨터(410)에 관하여 기술한 구성 요소들 중 다수 또는 모든 구성 요소를 포함한다. 도 4에 도시된 논리적 접속은 하나 이상의 근거리 네트워크(LAN; 471) 및 하나 이상의 광역 네트워크(WAN; 473)를 포함하지만, 다른 네트워크들도 포함할 수 있다. 이러한 네트워킹 환경은 사무실, 기업 광역 컴퓨터 네트워크(enterprise-wide computer network), 인트라넷 및 인터넷에서 일반적이다.
- [0052] LAN 네트워크 환경에서 사용되는 경우, 컴퓨터(410)는 네트워크 인터페이스 또는 어댑터(470)를 통해 LAN(471)에 접속된다. WAN 네트워크 환경에서 사용되는 경우, 컴퓨터(410)는 일반적으로 인터넷 등의 WAN(473)을 통해 통신을 구축하기 위한 모뎀(472) 또는 기타 수단을 포함한다. 내장형 또는 외장형일 수 있는 모뎀(472)은 사용자 입력 인터페이스(460) 또는 기타 적절한 메커니즘을 통해 시스템 버스(421)에 연결될 수 있다. 이를테면 인터페이스 및 안테나를 포함하는 무선 네트워킹 컴포넌트(474)가 액세스 포인트 또는 피어 컴퓨터와 같은 적절한 장치를 통해 WAN 또는 LAN에 연결될 수 있다. 네트워킹된 환경에서, 컴퓨터(410)에 관하여 도시된 프로그램 모듈들 또는 그 부분들은 원격 메모리 저장 장치에 저장될 수 있다. 제한이 아니라 예로서, 도 4는 원격 애플리케이션 프로그램(485)을 메모리 장치(481)에 상주하는 것으로서 도시한다. 도시된 네트워크 접속은 예시적이며, 컴퓨터들 사이에 통신 링크를 구축하는 다른 수단들이 사용될 수 있다는 것을 알 수 있다.
- [0053] (예를 들어, 콘텐츠의 보조 표시를 위한) 보조 서브시스템(499)이 사용자 인터페이스(460)를 통해 연결되어, 컴퓨터 시스템의 주요 부분들이 저전력 상태에 있는 경우에도, 프로그램 콘텐츠, 시스템 상태 및 이벤트 통지와 같은 데이터가 사용자에게 제공되게 할 수 있다. 보조 서브시스템(499)은 모뎀(472) 및/또는 네트워크 인터페이스(470)에 연결되어, 주요 처리 유닛(420)이 저전력 상태에 있는 동안에 그러한 시스템들 사이의 통신을 가능하게 할 수 있다.
- [0054] 결론
- [0055] 본 발명은 다양한 변경들 및 대안적인 구성들이 가능하지만, 본 발명의 소정의 실시예들이 도면들에 도시되고, 위에서 상세히 설명되었다. 그러나, 본 발명을 개시된 특정 형태들로 한정하는 것을 의도하는 것이 아니라, 본 발명의 사상 및 범위 내에 속하는 모든 변경들, 대안적인 구성들 및 균등물들을 포괄하는 것을 의도한다는 것을 이해해야 한다.

도면

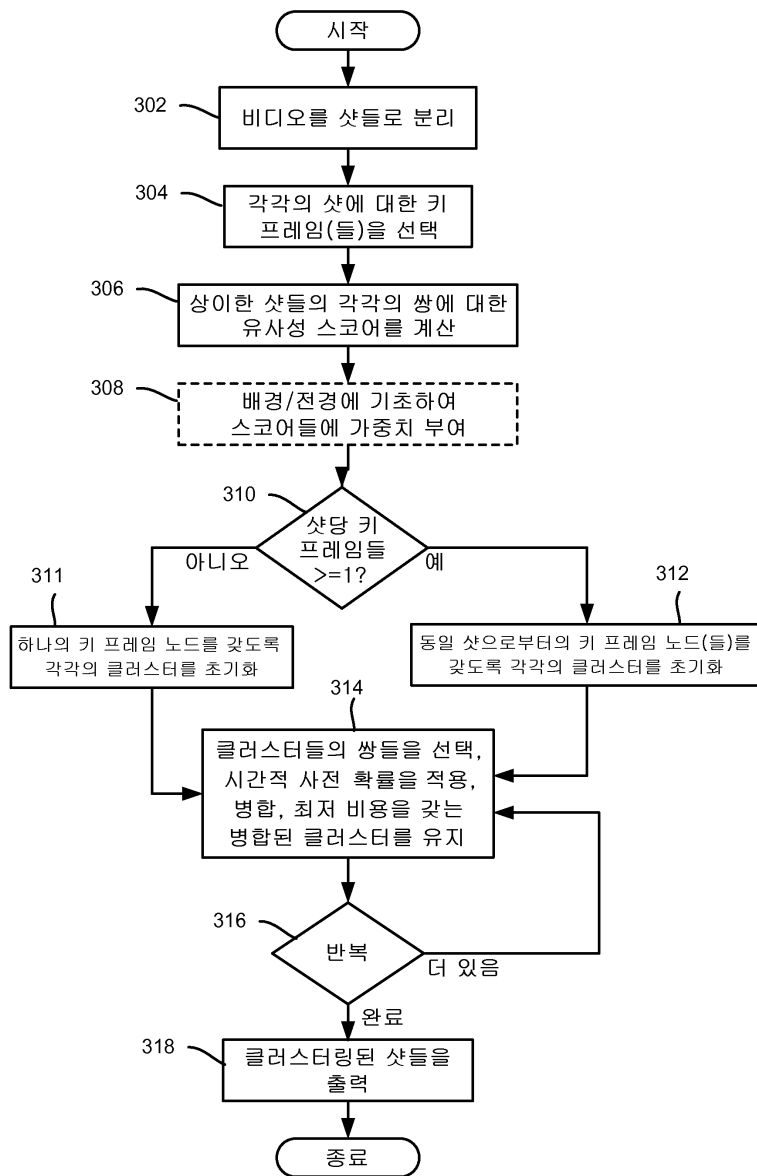
도면1



도면2



도면3



도면4

